

**KLASIFIKASI PENYAKIT DIABETES MELLITUS DENGAN MENGGUNAKAN
PERBANDINGAN ALGORITMA J48 DAN RANDOM FOREST (STUDI KASUS :
RUMAH SAKIT MUHAMMADIYAH LAMONGAN)**

Skripsi

Diajukan Untuk Memenuhi
Persyaratan Guna Meraih Gelar Sarjana Strata 1
Teknik Informatika Universitas Muhammadiyah Malang



SHAUMI RIZQI MAZIIDA
(201310370311083)

**JURUSAN TEKNIK INFORMATIKA
FAKULTAS TEKNIK
UNIVERSITAS MUHAMMADIYAH MALANG**

2018

LEMBAR PERSETUJUAN

Skripsi dengan Judul :

**KLASIFIKASI PENYAKIT DIABETES MELLITUS DENGAN
MENGUNAKAN PERBANDINGAN ALGORITMA J48 DAN
RANDOM FOREST (STUDI KASUS : RUMAH SAKIT
MUHAMMADIYAH LAMONGAN)**

Disusun Oleh :

SHAUMI RIZQI MAZIIDA


(201310370311083)

Sebagai persyaratan Guna Meraih Gelar Sarjana Strata 1

Teknik Informatika Universitas Muhammadiyah Malang

Menyetujui,

Dosen Pembimbing I



Setio Basuki, S.T., M.T.
NIP. 108.0907.0477

Dosen Pembimbing II



Agus Eko Minarno, M.Kom.
NIP.108.1410.0540



LEMBAR PENGESAHAN
KLASIFIKASI PENYAKIT DIABETES MELLITUS DENGAN
MENGGUNAKAN PERBANDINGAN ALGORITMA J48 DAN
RANDOM FOREST (STUDI KASUS RUMAH SAKIT
MUHAMMADIYAH LAMONGAN)

TUGAS AKHIR

Sebagai Persyaratan Guna meraih Gelar Sarjana Strata 1
Teknik Informatika Universitas Muhammadiyah Malang

Disusun Oleh :


SHAUMI RIZQI MAZHIDA

(201310370311083)

Tugas Akhir ini telah diuji dan dinyatakan lulus melalui sidang majelis penguji
pada tanggal April 2018


Menyetujui,

Penguji I,


Yuda Munarko, S.Kom., M.Sc.

NIDN: 0706077902


Penguji II,


Vinna Rahmayanti S.N

NIP:180306071990

Mengetahui,

Ketua Jurusan Teknik Informatika


Gita Indah M., S.T., M.Kom.

NIP.108.0611.0442

LEMBAR PERNYATAAN

Yang bertanda tangan di bawah ini :

NAMA : SHAUMI RIZQI MAZIIDA

NIM : 201310370311083

FAK./JUR. : TEKNIK/INFORMATIKA

Dengan ini saya menyatakan bahwa Tugas Akhir dengan judul “**KLASIFIKASI PENYAKIT DIABETES MELLITUS DENGAN MENGGUNAKAN PERBANDINGAN ALGORITMA J48 DAN RANDOM FOREST (STUDI KASUS : RUMAH SAKIT MUHAMMADIYAH LAMONGAN)**” beserta seluruh isinya adalah karya saya sendiri bukan merupakan karya tulis orang lain, baik sebagian maupun seluruhnya, kecuali dalam bentuk kutipan yang telah disebutkan sumbernya.

Demikian surat pernyataan ini saya buat dengan sebenar-benarnya. Apabila kemudian ditemukan adanya pelanggaran terhadap etika keilmuan dalam karya saya ini, atau ada klaim dari pihak lain terhadap keaslian karya saya ini maka saya siap menanggung segala bentuk resiko /sanksi yang berlaku.

Mengetahui,
Dosen Pembimbing



Setio Basuki, S.T., M.T.
NIP. 108.0907.0477

Malang, April 2018
Yang Membuat Pernyataan



Shaumi Rizqi Maziida

LEMBAR PERNYATAAN

Yang bertanda tangan di bawah ini :

NAMA : SHAUMI RIZQI MAZIIDA

NIM : 201310370311083

FAK./JUR. : TEKNIK/INFORMATIKA

Dengan ini saya menyatakan bahwa Tugas Akhir dengan judul **“KLASIFIKASI PENYAKIT DIABETES MELLITUS DENGAN MENGGUNAKAN PERBANDINGAN ALGORITMA J48 DAN RANDOM FOREST (STUDI KASUS : RUMAH SAKIT MUHAMMADIYAH LAMONGAN)”** beserta seluruh isinya adalah karya saya sendiri bukan merupakan karya tulis orang lain, baik sebagian maupun seluruhnya, kecuali dalam bentuk kutipan yang telah disebutkan sumbernya.

Demikian surat pernyataan ini saya buat dengan sebenar-benarnya. Apabila kemudian ditemukan adanya pelanggaran terhadap etika keilmuan dalam karya saya ini, atau ada klaim dari pihak lain terhadap keaslian karya saya ini maka saya siap menanggung segala bentuk resiko /sanksi yang berlaku.

Mengetahui,
Dosen Pembimbing

Malang, April 2018
Yang Membuat Pernyataan

Setio Basuki, S.T., M.T.
NIP. 108.0907.0477

Shaumi Rizqi Maziida

LEMBAR PERSEMBAHAN

Melalui kesempatan ini penulis mengucapkan terimakasih kepada pihak-pihak yang telah banyak member dukungan selama proses penulisan sehingga penulis bisa menyelesaikan tugas akhir ini.

Penulis menyampaikan ucapan terima kasih yang sebesar-besarnya kepada :

1. Allah SubhanahuWaTa'ala yang maha member petunjuk serta kemudahan.
2. Bapak Setio Basuki,ST.,MT dan Bapak Agus Eko Minarno, M.Kom selaku Dosen Pembimbing yang selalu memberi arahan dan menyempatkan waktu untuk membimbing saya dalam menyelesaikan tugas akhir.
3. Kedua orang tua saya, yang selalu member do'a dan dukungan baik secara moral dan materi.
4. Sahabat-sahabat saya Lutfi Meidina, Khoirir Rosikin, Andrik Setiawan, Mairisa Anggraini, Oktavianas dan Muhammad Ubaidillah yang selalu membantu dalam menyelesaikan tugas akhir.
5. Pihak dosen beserta staff TU Jurusan Teknik Informatika UMM.
6. Kakak Tingkat 2012 yang banyak membantu dalam pengerjaan proposal tugas akhir.
7. Teman-temak TIF 2013.
8. Darmawan Izaad Ibrahim Asari yang banyak membantu mulai dari pengerjaan proposal sampai dengan tugas akhir ini selesai.

KATA PENGANTAR



Assalamu'alaikum Wr. Wb.

Segala puji bagi Allah SWT, yang telah memberikan Rahmat dan Karunianya, sehingga penulis dapat menyelesaikan skripsi yang berjudul: **“KLASIFIKASI PENYAKIT DIABETES MELLITUS DENGAN MENGGUNAKAN PERBANDINGAN ALGORITMA J48 DAN RANDOM FOREST (STUDI KASUS : RUMAH SAKIT MUHAMMADIYAH LAMONGAN)”**

Skripsi ini diajukan untuk memenuhi salah satu syarat dalam memperoleh gelar sarjana pada program Sarjana Fakultas Teknik Universitas Muhammadiyah Malang. Skripsi ini merupakan salah satu syarat studi yang harus ditempuh oleh seluruh mahasiswa Universitas Muhammadiyah Malang, guna menyelesaikan akhir studi pada jenjang program Strata 1.

Penulis menyadari masih terdapat banyak kekurangan dan keterbatasan didalam penulisan skripsi ini. Oleh sebab itu penulis mengharapkan adanya saran yang dapat membantu agar tulisan ini dapat berguna untuk perkembangan ilmu pengetahuan teknologi informasi kedepannya.

Malang, Maret 2018

Penulis

Shaumi Rizqi Maziida

ABSTRAK

Diabetes Mellitus (DM) merupakan penyakit yang terjadi akibat kadar glukosa yang tinggi dalam darah, dalam penelitian ini data yang didapatkan dari Rumah Sakit Muhammadiyah Lamongan terdapat 2.021 data original yang didalamnya terdapat 5 jenis penyakit DM atau target kelas. Kemudian pada data tersebut dilakukan pemrosesan berupa normalisasi dan penanganan missing values untuk selanjutnya dilakukan proses klasifikasi dan pengujian. Hasil pengujian algoritma Random Forest memiliki hasil akurasi yang lebih unggul dibanding dengan J48 pada pengujian data original dan data yang telah melalui proses normalisasi dan penanganan missing values. Pada data original pengujian menggunakan confusion matrix pada algoritma Random Forest memiliki akurasi 98,51% sedangkan J48 95,05%. Dan pada data yang telah diproses dengan normalisasi dan penanganan missing values algoritma Random Forest memiliki akurasi 97,52% sedangkan J48 94,76%. Untuk perbandingan data yang digunakan data original memiliki akurasi yang lebih baik dibandingkan dengan data yang telah diproses.

ABSTRACT

Diabetes Mellitus (DM) is a disease that occurs due to high glucose levels in the blood. in this research, data obtained from Muhammadiyah Lamongan Hospital is 2.021 original data and there are 5 types of diabetes mellitus or target class therein. Then the data is processed using normalization method and handling missing values for classification and testing process. Random Forest algorithm test results have superior accuracy results compared with J48 in the original data test and data normalization and handling of missing values. In the original data, the test with confusion matrix on the Random Forest algorithm has an accuracy of 98.51% and J48 is 95.05%. And on the data has been through the process of normalization and handling of missing values Random Forest algorithm accuracy is 97.52% and J48 is 94.,76%. The original data test results have better accuracy than the data that has been processed.



DAFTAR ISI

LEMBAR PERSETUJUAN	i
LEMBAR PENGESAHAN	ii
LEMBAR PERNYATAAN	iv
LEMBAR PERSEMBAHAN	v
KATA PENGANTAR	vi
ABSTRAK	vii
ABSTRACT	viii
DAFTAR ISI.....	ix
DAFTAR GAMBAR.....	xii
DAFTAR TABEL	xiii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	2
1.3 Tujuan Penelitian.....	3
1.4 Batasan Masalah	3
1.5 Metodologi	3
1.5.1 Studi Pustaka.....	3
1.5.2 Pengumpulan Data	4
1.5.3 Desain Sistem.....	4
1.5.4 Implementasi Metode.....	4
1.5.4 pengujian	4
1.5.4 Pembuatan Laporan.....	4
1.6 Sistematika Penulisan	4
BAB II LANDASAN TEORI	6
2.1 Diabetes Mellitus	6

2.2	Data Mining	6
2.3	Klasifikasi	7
2.4	Pembelajaran Mesin	7
2.4.1	Definisi	7
2.4.2	Sejarah Machine Learning	8
2.4.3	Penerapan Machine Learning	8
2.5	Pengolahan Data	8
2.5.1	Pengumpulan Data	8
2.5.2	Data Preprocessing	8
2.6	Algoritma Decision Tree J48	9
2.7	Algoritma Random Forest	11
2.7	Pengujian	12
2.7.1	Pengujian Pengelolaan Data	12
2.4.2	Pengujian Confusion Matrix	13
BAB III ANALISA DAN PERANCANGAN SISTEM		15
3.1	Analisa Masalah	15
3.2	Analisa Sistem	15
3.2.1	Alur Sistem	15
3.3	Pengumpulan Data	17
3.4	Analisa <i>Pre-processing</i>	18
3.4.1	Normalisasi	18
3.4.2	Missing Valuses	20
3.5	Metode J48	21
3.6	Metode Random Forest	21
3.7	Perancangan Pengujian	22
3.7.1	Perancangan Pengujian Dalam Pengelolaan Data	22

3.7.2	Pengujian Confusion Matrix	22
3.8	Perancangan Antar Muka (<i>Interface</i>)	24
BAB IV	IMPLEMENTASI DAN PENGUJIAN SISTEM	27
4.1	Implementasi	27
4.1.1	Lingkungan Implementasi.....	27
4.1.2	Batasan Implementasi	27
4.2	Implementasi Interface.....	27
4.2.1	Implementasi Tampilan Awal	28
4.2.2	Implementasi tampilan Untuk Normalisasi.....	28
4.2.3	Implementasi Tampilan Untuk Data Cleaning.....	29
4.2.4	Implementasi Tampilan Untuk Pengujian J48	30
4.2.5	Implementasi Tampilan Untuk Pengujian Random Forest	31
4.3	Hasil Pengujian.....	31
4.3.1	Pengujian Dengan Data Original	32
4.3.2	Pengujian Dengan Data Yang Sudah Diproses	34
4.3	Analisis Hasil Pengujian.....	38
BAB V	PENUTUP.....	40
5.1	Kesimpulan	40
5.2	Saran	40
DAFTAR PUSTAKA		41
Lampiran		43
	Lampiran Data Original	43

DAFTAR GAMBAR

Gambar 2.1 Tahapan Klasifikasi.....	7
Gambar 2.2 Proses Model Menjadi Tree	10
Gambar 2.3 Proses Klasifikasi Random Forest.....	12
Gambar 3.1 Alur Sistem Klasifikasi Diabetes Mellitus	16
Gambar 3.2 Alur Data Preprocessing.....	18
Gambar 3.3 Alur Metode J48.....	21
Gambar 3.4 Alur Metode Random Forest.....	21
Gambar 3.5 Perancangan Antar Muka Untuk Proses Normalisasi	25
Gambar 3.5 Perancangan Antar Muka Untuk Proses Missing Values.....	25
Gambar 3.6 Perancangan Antar Muka Untuk Proses Klasifikasi	26
Gambar 4.1 Implementasi Tampilan Awal	28
Gambar 4.2 Implementasi Tampilan Untuk Normalisasi	28
Gambar 4.3 Implementasi Tampilan Data Cleaning.....	29
Gambar 4.4 Implementasi Tampilan Pengujian J48	30
Gambar 4.5 Implementasi Tampilan Pengujian Random Forest	31
Gambar 4.6 Out Of Bag Pengujian Data Original	34
Gambar 4.7 Out Of Bag Pengujian Data Olahan	38

DAFTAR TABEL

Tabel 3.1 Tabel Dataset Penderita Diabetes Mellitus	17
Tabel 3.2 Tabel Data Training (yang belum dinormalisasi)	18
Tabel 3.3 Tabel Data Training (yang dinormalisasi)	19
Tabel 3.4 Tabel Data Testing (yang belum dinormalisasi)	19
Tabel 3.5 Tabel Data Testing (yang dinormalisasi)	20
Tabel 3.6 Tabel Pengelolaan Missing Value.....	20
Tabel 3.7 Contoh Perancangan Pengujian Klasifikasi	23
Tabel 3.8 Tabel Skenario Pengujian Confusion Matrix.....	23
Tabel 4.1 Tabel Pengujian Presisi dan Recall Metode J48	32
Tabel 4.2 Tabel Pengujian Presisi dan Recall Metode Random Forest	33
Tabel 4.3 Tabel Pengujian Akurasi Dengan Data Original.....	33
Tabel 4.4 Tabel Data Normalisasi.....	34
Tabel 4.5 Tabel Data Cleaning.....	35
Tabel 4.6 Tabel Pengujian Presision dan Recall Metode J48 dengan Data Olahan ..	36
Tabel 4.7 Tabel Pengujian Presision dan Recall Metode Random Forest dengan Data Olahan.....	37
Tabel 4.8 Tabel Pengujian Akurasi Dengan Data Yang Sudah Diproses	37
Tabel 4.9 Analisis Hasil Pengujian Kedua Data	39

DAFTAR PUSTAKA

- [1] Lesmana, I Putu Dody, 2012. Pengembangan Decision Tree J48 Untuk Diagnosis Penyakit Diabetes Mellitus. Jember. Jember, Jawa Timur. Politeknik Negeri Jember.
- [2] Nidhomuddin, Bambang Widjanarko Otok, 2015. Random Forest dan Multivariate Adaptive Regression Spline (MARS) Binary Response Untuk Klasifikasi Penderita HIV/AIDS di Surabaya. Surabaya, Jawa Timur, Jurusan Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Institut Teknologi Sepuluh November Surabaya.
- [3] Sambodo, Katmoko Ari. Rahayu, Mulia Inda. Indriasari, Novie. M. Natsir, 2014. Klasifikasi Hutan-Non Hutan Data Alos Palsar Menggunakan Metode Random Forest. Pusat Teknologi dan Data Penginderaan Jauh, LAPAN.
- [4] Kementerian Kesehatan RI, 2014. Waspada Diabetes, Jakarta Selatan, Pusat Data dan Informasi.
- [5] Asriningtias, Yuli. Mardhiyah, Rodhyah. 2014. Aplikasi Data Mining Untuk Menampilkan Tingkat Kelulusan Mahasiswa. Yogyakarta. Universitas Teknologi Yogyakarta.
- [6] Classification vs Clustering, "BAB 2 LANDASAN TEORI 2.1.4 Classification vs Clustering" pp. 11-38, 2003 [Online]. Available: <http://library.binus.ac.id/eColls/eThesisdoc/Bab2HTML/2013100584SIBab2003/body.html>
- [7] Purnamasari, Detty. Henharta, Jonathan. Sasmita, Yoga Perdana. Ihsani, Fuji. Wicaksana, I Wayan Simri. 2013. Get Easy Using Weka. Jakarta Timur. Dapur Buku.

- [8] Elvitaria, Luluk. Havenda, Muhammad. 2017. Memprediksi Tingkat Peminat Ekstrakurikuler Pada Siswa SMK Analisis Kesehatan Abdurrah Menggunakan Algoritma C4.5. Pekanbaru. Riau. Universitas Abdurrah.
- [9] Nurfaizah. Imron, Muhammad. Perdanawanti, Linda. 2017. Algoritma Decision Tree, K-Nearest, dan Zero-R Pada Kinerja Akademik. Purwokerto. STMIK AMIKOM Purwokerto.
- [10] Kumara, rendragraha. Supriyanto, Catur. Klasifikasi Data Mining Untuk Penerimaan Calon Pegawai Negeri Sipil 2014 Menggunakan Algoritma Decision Tree C4.5. Semarang. Universitas Dian Nuswantoro Semarang.
- [11] Menarianti, Ika. 2015. Klasifikasi Data Mining Dalam Menentukan Pemberian Kredit Bagi Nasabah Koperasi. Semarang. Universitas PGRI Semarang.

